AD-A097 688 MINNESOTA UNIV MINNEAPOLIS COMPUTERIZED ADAPTIVE TES--ETC F/G 5/9
EFFECTS OF IMMEDIATE FEEDBACK AND PACING OF ITEM PRESENTATION O--ETC(U)
FEB 81 M F JOHNSON, D J WEISS, J S PRESTWOOD N00014-76-C-0243
UNCLASSIFIED RR-81-2 NL

END
DATE FILMED
5 81
DTIC

12

LEVEL II

Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing

Marilyn F. Johnson
David J. Weiss
J. Steven Prestwood

RESEARCH REPORT 81-2
FEBRUARY 1981

COMPUTERIZED ADAPTIVE TESTING LABORATORY
PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Navy Personnel Research and Development Center, the Air Force Human Resources Laboratory, the Air Force Office of Scientific Research, the Army Research Institute, and the Office of Naval Research, and monitored by the Office of Naval Research.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Research Report 81-2	2. GOVT ACCESSION NO.: AD-A097 688
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing,	5. TYPE OF REPORT & PERIOD COVERED: Technical Report,
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Marilyn F. Johnson, David J. Weiss and J. Stephen Prestwood	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0243; N00014-79-C-0172
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, University of Minnesota, Minneapolis, MN 55455	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: P.E.: 6115N Proj.: RR042-04; T.A.: RR042-04-01; W.U.: NR 150-433, NR 150-382
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research, Arlington, Virginia 22217	12. REPORT DATE: February 1981
	13. NUMBER OF PAGES: 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. ~~Reproduction in whole or in part is permitted for any purpose of the United States Government.~~

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by funds from the Navy Personnel Research and Development Center, the Air Force Human Resources Laboratory, the Air Force Office of Scientific Research, the Army Research Institute, and the Office of Naval Research, and monitored by the Office of Naval Research.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Testing	Programmed Testing	Feedback
Adaptive Testing	Individualized Testing	Psychological Effects
Computerized Testing	Branched Testing	Response Latency
Tailored Testing	Response-Contingent Testing	Pacing of Item Presentation
Ability Testing	Latent-Trait Theory	Knowledge of Results
		Test Reactions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

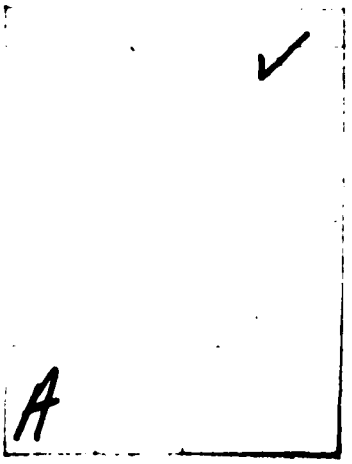
The study investigated the joint effects of knowledge of results (KR or no-KR), pacing of item presentation (computer or self-pacing), and type of testing strategy (50-item peaked conventional, variable-length stradaptive, or 50-item fixed-length stradaptive test) on ability test performance, test item response latency, information, and psychological reactions to testing. The psychological reactions to testing were obtained from Likert-type items that assessed test-taking anxiety, motivation, perception of difficulty, and reac-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

tions to knowledge of results. Data were obtained from 447 college students randomly assigned to one of the 12 experimental conditions. The results indicated that there were no effects on ability estimates due to knowledge of results, testing strategy, or pacing of item presentation. Although average latencies were greater on the stradaptive tests than on the conventional test, the overall testing time was not substantially longer on the adaptive tests and may have been a function of differences in test difficulty. Analysis of information values indicated higher levels of information on the stradaptive tests than on the conventional test. There was no statistically significant main effect for any of the three experimental conditions when test anxiety or test-taking motivation were the dependent variables, although there were some significant interaction effects. These results indicate that testing conditions may interact in a complex way to determine psychological reactions to the testing environment. The interactions do suggest, however, a somewhat consistent standardizing effect of KR on test anxiety and test-taking motivation. This standardizing effect of KR showed that approximately equal levels of motivation and anxiety were reported under the various testing conditions when KR was provided, but that mean levels of these variables were substantially different when KR was not provided. Consistent with theoretical expectations, the conventional test was perceived as being either too easy or too difficult, whereas the adaptive tests were perceived more often as being of appropriate difficulty. The results concerning the effects of KR on test performance, motivation, and anxiety found in this study were contrary to earlier reported findings; and differences in the studies are delineated. Recommendations are made concerning the control of specific testing conditions, such as difficulty of the test and ability level of the examinee population, as well as suggestions for the further analysis of the standardizing effect of KR.



CONTENTS

Technical Editor: Barbara Leslie Camm

Effects of Immediate Feedback and Pacing of Item Presentation on Ability Test Performance and Psychological Reactions to Testing

The motivation to perform well on an ability test has been suggested as a significant factor affecting test performance (Cronbach, 1970). Some researchers (e.g., Bayroff, 1964; Betz, 1975; Betz & Weiss, 1976b; Ferguson & Hsu, 1971; Strang & Rust, 1973; Zontine, Richards, & Strang, 1972) have hypothesized that immediate feedback or knowledge of results (KR) may increase motivation to perform well. Others have suggested mechanisms by which KR affects behavior. Locke, Cartledge, and Koeppel (1968) have offered an explanation for the way in which motivation is affected by KR through goal-setting behavior. They have indicated that KR may mediate test-taking behavior if an examinee makes an evaluation of performance in response to receiving KR and adjusts his/her subsequent level of effort. In such a process the examinee sets a goal, and the intent to achieve that goal alters test-taking behavior. In this view, KR without goal-setting does not influence test-taking behavior. Such an explanation presupposes that goals and intentions influence behavior.

Another way in which KR is hypothesized to affect test-taking behavior is through increasing or decreasing test anxiety (Liebert & Morris, 1967; Morris & Fulmer, 1976). Negative or failure feedback is hypothesized to increase anxiety, and positive feedback is hypothesized to decrease anxiety. Failure feedback would tend to increase expectancy of poor performance and thus tend to increase worry or concern about test performance. It is suggested, however, that failure feedback may have a facilitative or motivating effect on low-anxiety (high-ability) students. Examinees with expectations of good performance, in general, would be less anxious about test performance. Thus, according to this conceptualization, test anxiety varies inversely with expectancy of performance. In addition, Liebert and Morris (1967) and Morris and Fulmer (1976) posit that test anxiety has a detrimental effect on test performance. They also state that feedback affects test performance because of the certainty an examinee attaches to judgments of performance level. Thus, it is not only expectancy of test performance that is affected by feedback but also certainty. Two examinees with the same expectancy of performance may differ with respect to the certainty of their judgment of performance because one received feedback and the other did not. The examinee with the greater certainty, which is insured by providing accurate feedback, will have less anxiety than the examinee attaching less certainty to his/her judgment of test performance, even though he/she expects to do more poorly.

Although two seemingly different mechanisms have been hypothesized to account for the way in which KR mediates test-taking behavior, the motivation variable discussed by Locke et al. (1968) has some similarity to the anxiety variable proposed by Morris and Fulmer (1976). The mechanism that motivates an examinee to try harder on a test in response to negative KR may be termed facil-

itating anxiety (Betz & Weiss, 1976b; Mandler & Sarason, 1952). Although Morris and Fulmer (1976) stated that negative KR makes examinees more anxious and that anxiety detrimentally affects test performance, it is possible that such anxiety may serve to improve or to facilitate test performance. Certain groups, for example, high-ability college students, may respond more often to negative KR by trying harder on the test than would lower ability students (Betz & Weiss, 1976a).

Immediate feedback of test performance, although difficult to provide when administering a paper-and-pencil test, is a relatively simple procedure when tests of ability or achievement are administered by computer. For this reason, adaptive testing research has facilitated investigation of the effects KR might have on test performance and on psychological reactions, such as motivation and anxiety, toward testing. Betz and Weiss (1976a, 1976b) studied the effects of KR on high- and low-ability students taking either a 50-item conventional test or a stradaptive test. They found that mean test performance as measured by maximum likelihood ability estimates was higher when KR was provided than when it was not (Betz & Weiss, 1976a). They also found that mean number-correct score was higher under KR than under no-KR conditions for students taking a conventional (i.e., nonadaptive) test. Betz and Weiss (1976b) found no main effect due to KR on measures of motivation and anxiety as assessed by a posttest questionnaire of Likert-type items. However, a significant interaction between KR and ability indicated that among high-ability students those receiving KR reported a higher level of motivation than those not receiving KR, whereas low-ability students receiving KR reported a lower level of motivation than those not receiving KR.

Prestwood and Weiss (1978) studied the effects of KR and test difficulty on test performance and on psychological reactions to testing in high-ability students. They found that, similar to the results reported by Betz and Weiss (1976b), there was no difference in anxiety due to provision of KR; but among high-ability college students, motivation was higher when KR was provided than when it was not. In addition, a marginally significant ($p = .054$) interaction between KR and test difficulty factors on maximum likelihood ability estimates indicated that when KR was provided, the mean ability estimate was highest on the most difficult tests and lowest on the least difficult tests. The mean ability estimate for students in the no-KR conditions was highest on the least difficult tests and lowest on the most difficult tests.

Purpose

In the studies presented above (Betz & Weiss, 1976a, 1976b; Prestwood & Weiss, 1978), which were designed to assess the effects of KR, the provision of KR was confounded with pacing of item presentation. That is, the rate or pacing of item presentation in the KR condition differed from the pacing in the no-KR condition. In those studies, tests in the KR condition were essentially self-paced; after receiving appropriate feedback following an item response, students typed the letter "P" (for proceed) on the terminal keyboard and pressed the "Return" key in order to initiate the presentation of the next test item. In the no-KR condition, tests were computer-paced, i.e., students not receiving KR were automatically presented with the next test item immediately following their response to each item.

The present study was designed to separately examine the effects of KR and of computer- versus self-pacing of item presentation in order to determine the unconfounded contribution of each to the results reported in the research cited. Since the effects of KR and pacing might differ depending on whether an adaptive or a conventional test was administered, the study investigated the joint effects of KR, pacing of item presentation, and type of testing strategy on ability test performance and on psychological reactions to testing.

METHOD

Procedure

Subjects

The 447 subjects who participated in this experiment were students drawn from an introductory psychology course at the University of Minnesota. The students were volunteers who received points that counted toward their final course grade in return for their participation in the experiment.

Test Administration

Students were assigned sequentially to testing conditions in which they took a test at an individual cathode-ray terminal (CRT). Each terminal was connected to a Hewlett-Packard real-time computer system. A test proctor was present in the testing room to provide assistance to any examinee during testing. Students were assured that they could take as much time as necessary to complete the test.

Prior to actual testing, instructional screens explaining the operation of the CRTs were displayed. After students reviewed the test instructions and responded to a number of identification and demographic questions, the experimental test was administered. Each experimental ability test was composed of five-alternative multiple-choice vocabulary questions, which students answered by typing a number on the CRT keyboard that corresponded to the chosen alternative. Following the experimental test, examinees not receiving feedback recorded their reactions to the test by responding to 18 Likert-type questions. Students receiving feedback responded to the same 18 questions as well as to 8 additional questions concerning their reactions toward feedback.

Design

Independent Variables

This study analyzed three independent variables in a $3 \times 2 \times 2$ completely crossed design. One factor was ability test strategy. The three levels of this factor were (1) a 50-item peaked conventional test, (2) a variable-length stradaptive test (Weiss, 1973), and (3) a fixed-length (50-item) stradaptive test. The second factor was immediate knowledge of results (KR); there were two levels of this factor: (1) KR and (2) no-KR. The third factor was pacing of item presentation: Items were either (1) computer-paced or (2) self-paced.

Students in the KR conditions were informed by the computer immediately

after their response to a test question whether it was correct or incorrect; if it was incorrect, the correct alternative was given. In the KR computer-paced conditions this was followed by a 4-second delay until the next question was presented. In the no-KR computer-paced conditions there was no delay between item response and presentation of the next item. Under self-pacing conditions, either with or without KR, students could pace the rate of item presentation: After responding to an item, the next item could be presented by typing "P" (for proceed) and pressing the "Return" key.

Dependent Variables

Ability estimates. A major dependent variable of interest was test performance of the examinee, which was estimated in three ways. Performance on the stradaptive and conventional tests was assessed by maximum likelihood ability estimates computed for each person by employing the likelihood equation for Birnbaum's (1968) three-parameter logistic model. A second ability measure was proportion-correct scores, which were computed for students who took the conventional test. The proportion-correct score, which is the ratio of number of items answered correctly to total number of items administered, is an inappropriate measure of ability when a test is adapted to an individual's level of ability (Weiss, 1973, 1974). For this reason, proportion-correct scores were not computed in stradaptive testing conditions. The third ability measure, used only for the stradaptive tests, was the mean difficulty correct score, which was found in previous stradaptive testing research to be a valid (Thompson & Weiss, 1980) and reliable (Vale & Weiss, 1975a, 1975b) approach to ability estimation in stradaptive tests. The mean difficulty correct score was computed by averaging the normal ogive difficulty parameters of the items answered correctly on the stradaptive test by each individual.

Response latency. Mean latency of response was calculated for each individual. Measured in seconds, this value represents the average time it took an individual to read and respond to an item. However, since the length of each item was quite similar, latency would serve as a rough indication of the "processing time" required by the individual to answer an item. The mean latency measured was mean latency over all items administered, in order to determine whether testing conditions affected processing time.

Information. Information is an index of precision of measurement (Bejar & Weiss, 1979; Bejar, Weiss, & Gialluca, 1977). Although information is similar in function to reliability, it differs in that information values are appropriate in describing precision at any level of the trait continuum. Thus, test information can be used to evaluate testing strategies (e.g., Bejar et al., 1977; Betz & Weiss, 1974, 1975; McBride & Weiss, 1976; Vale, 1975; Vale & Weiss, 1975b, 1977). For example, testing strategies with high information values over all trait levels are to be preferred to tests with either low or peaked information curves. In this study comparisons between testing conditions were based on response pattern information values derived from the second derivative of the log-likelihood function evaluated at each individual's final ability estimate (Bejar & Weiss, 1979). These response pattern information values were calculated for each person in every experimental condition.

Psychological reactions. Psychological reactions to the testing conditions were also of interest in this study. Measures of reported anxiety, motivation,

and perception of test difficulty were obtained from Likert-type items. In addition, those students in the KR conditions provided data on their reactions to KR. Scales used to measure these variables were those used by Prestwood and Weiss (1978). A total of 26 questions were administered to the students in the KR conditions; those in the no-KR conditions responded to 18 psychological reactions questions; these questions were administered immediately following the experimental tests. The questions and their order of administration are shown in Table 7 (Motivation), Table 9 (Anxiety), Table 11 (Difficulty Perception), and Table 13 (KR Reaction); Table 14 shows eight additional questions administered that were not included in these four scales.

Test Construction

Item Pool

The item pool from which the conventional and stradaptive test items were drawn consisted of 569 five-alternative multiple-choice vocabulary items. Item response function (IRF, or item characteristic curve) parameter estimates were obtained from samples of the college student population, according to the procedure described by Prestwood and Weiss (1977, Appendix A). Each item had associated with it a normal ogive discrimination (a) and difficulty (b) parameter estimate. The "guessing" (c) parameter of each of the five-alternative items was assumed to be .20.

Conventional Test

The peaked 50-item conventional test was composed of items whose difficulty parameters centered around the ability level of the student population. Fifty items were chosen so that the mean difficulty of the items matched the estimated ability level of students taking the test and so that normal ogive discrimination parameters were greater than or equal to $a = .40$ (see Appendix Table A). The mean of the difficulty parameters was $b = .02$, although the values varied from $b = -.355$ to $b = +.334$, with a standard deviation of .20. The mean of the discrimination values was $a = .88$, with values ranging from $a = .407$ to $a = 1.96$ and a standard deviation of .35.

Stradaptive Tests

Stradaptive testing required a stratified item pool (Weiss, 1973) with items grouped by difficulty (b) parameters into nine nonoverlapping strata. Within a stratum, items were arranged in descending order of their discrimination (a) parameter estimates. The number of items in each stratum varied, ranging from 16 items in Stratum 9 (the most difficult stratum) to 57 items in Stratum 7. Three hundred twenty-five items were selected from the total item pool so that no item with a discrimination parameter estimate less than $a = .30$ was included in the stradaptive item pool. Appendix Table B shows IRF parameter estimates for items in the stradaptive item pool.

The entry point or stratum level from which the first item was selected for administration was based on student-estimated college grade-point-average (GPA) level. Students with higher reported GPA received an item from a correspondingly difficult stratum, following the procedure used by Thompson and Weiss (1980, p. 5). Thereafter, items were selected according to an "up-one/down-one"

branching procedure. By this method, correct answers resulted in branching to the next more difficult stratum, whereas incorrect answers routed a testee to an item at the next easier stratum of items. At every point during testing, the most discriminating item of those remaining in a given stratum that had not already been administered to a given individual was selected as the next item to be administered.

Testing continued until 50 items had been administered in the fixed-length stradaptive test, but testing terminated in the variable-length stradaptive test when conditions set by the termination criterion were met. According to this criterion, testing was terminated when a stratum was identified at which a student responded to a series of items at chance level or below. Chance level was defined to be the reciprocal of the number of alternatives in the multiple-choice question. In this case, the multiple-choice questions each had five alternatives, so the chance level of responding was set as a proportion correct of .20 within a stratum. In order to implement this condition, however, a minimum of five questions within a stratum were required to be administered prior to termination. If the termination criterion was not reached by administration of the 75th item, testing was terminated at that point.

Data Analysis

Ability Estimates

Maximum likelihood. Maximum likelihood ability estimates were calculated for each examinee. These ability estimates were analyzed using a 3 × 2 × 2 completely crossed analysis of variance in which testing strategy, KR condition, and pacing of item presentation were independent variables. Means and standard deviations for each experimental treatment combination were also computed for this variable.

Proportion correct. The proportion of items answered correctly was computed for those students in the conventional test conditions. Means and standard deviations of this variable were calculated in all KR and pacing conditions. Proportion-correct scores also served as a dependent variable in a two-way analysis of variance in which KR and pacing of item presentation were independent variables within the conventional test conditions.

Mean difficulty correct. Within the stradaptive testing condition the mean difficulty correct score was analyzed in a 2 × 2 × 2 crossed analysis of variance in which each factor--stradaptive test condition (fixed vs. variable length), KR condition, and pacing condition--had two levels. Means and standard deviations for this dependent variable were also computed in all experimental treatment combinations.

Response Pattern Information

Response pattern information was computed for each individual at the last iteration of the maximum likelihood scoring of the individual's test response data and served as the dependent variable in a 3 × 2 × 2 analysis of variance. Means and standard deviations of this variable were computed for each combination of experimental conditions.

Response Latency

Mean response latency across all items administered, and scores on psychological reactions scales derived from the factor analysis, were dependent variables in univariate 3 × 2 × 2 analyses of variance in which testing strategy, KR condition, and pacing condition were independent variables. In addition, students in the KR conditions yielded scores on a KR Reaction Scale; these values were analyzed in a 3 × 2 analysis of variance. Means and standard deviations were also computed on all variables in the combined test strategy, pacing, and KR conditions.

Psychological Reactions

In order to further examine students' reactions to testing within experimental groups, the percentages of students who chose each response alternative in each psychological reactions question were calculated for the total group and for each experimental group. Chi-square tests of independence were computed to identify reactions to testing at the single question level which differed among the experimental conditions. Comparisons of item responses were made between three pairs of experimental conditions: KR versus no-KR condition, conventional versus stradaptive testing strategy, and self- versus computer-paced condition on non-KR items. For comparisons involving stradaptive and conventional test strategies, data for fixed-length and variable-length stradaptive testing strategies were combined. On KR Reaction Scale items, comparisons were made between testing strategies and pacing conditions.

Finally, to examine the nature of the relationships among the dependent variables, intercorrelations were computed among the dependent variables, and the internal consistency reliability of the psychological reactions scales was determined by Cronbach's alpha.

RESULTS

Ability Estimates

Maximum likelihood ability estimates. Table 1 shows the results of the three-way analysis of variance in which the effects of testing strategy, KR, and pacing of item presentation on maximum likelihood ability estimates were analyzed. Also shown are the means and standard deviations of the maximum likelihood ability estimates in each of the experimental groups and the number of subjects associated with them. As the results of the three-way analysis of variance show, there was no significant effect on maximum likelihood ability estimates due to testing strategy, KR condition, or pacing of item presentation; and there were no significant interactions.

Proportion-correct scores. The results of the two-way analysis of variance that analyzed the effects KR and pacing condition had on proportion-correct scores obtained in the conventional testing condition are presented in Table 2. Also shown are means and standard deviations of the proportion-correct scores in each of the experimental conditions; the analysis indicated that there was no significant effect of KR or pacing condition on proportion-correct scores, nor was there a significant interaction.

Table 1
Means and Standard Deviations of Maximum Likelihood Ability Estimates for Conventional and Stradaptive Tests With and Without KR in Computer- and Self-Paced Conditions, and Three-Way ANOVA Results

Test and KR Condition	Experimental Condition: Self-Paced N	Mean	SD	Computer-Paced N	Mean	SD	Combined Conditions N	Mean	SD
Conventional									
KR	39	-.13	1.22	39	-.32	1.44	78	-.23	1.33
No-KR	31	-.20	1.34	32	-.40	1.58	63	-.30	1.46
Stradaptive: Fixed Length									
KR	41	-.18	1.13	41	-.28	1.01	82	-.23	1.06
No-KR	33	-.51	.94	33	.09	1.05	66	-.21	1.04
Stradaptive: Variable Length									
KR	39	-.01	.88	40	-.09	1.05	79	-.05	.96
No-KR	33	-.17	1.04	33	-.11	1.13	66	-.14	1.11
Combined Groups									
Conventional	70	-.16	1.26	71	-.36	1.50	141	-.26	1.38
Stradaptive									
Fixed Length	74	-.32	1.06	74	-.11	1.04	148	-.22	1.05
Variable Length	72	-.08	.98	73	-.10	1.08	145	-.09	1.03
KR	119	-.11	1.08	120	-.23	1.17	239	-.17	1.13
No-KR	97	-.30	1.13	98	-.14	1.27	195	-.22	1.21
Total Group	216	-.19	1.11	218	-.19	1.22	434	-.19	1.16

Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	4	.64	.47	.760
Test	2	1.14	.84	.434
KR	1	.28	.20	.654
Pacing	1	.00	.00	.975
Two-Way Interactions	5	1.06	.78	.568
Test × KR	2	.12	.09	.914
Test × Pacing	2	1.48	1.09	.338
KR × Pacing	1	2.08	1.53	.217
Three-Way Interaction				
Test × KR × Pacing	2	1.26	.923	.398
Residual	422	1.36		
Total	433	1.35		

[a]Probability of error in rejecting null hypothesis.

Mean difficulty correct scores. Table 3 shows the three-way analysis of variance and descriptive statistics when mean difficulty correct scores were computed for items answered correctly. Although the KR × Pacing interaction approached significance ($p \leq .088$), there were no other significant sources of variance in the data.

Table 2
Means and Standard Deviations of Proportion-Correct Scores for Conventional Test With and Without KR in Computer- and Self-Paced Conditions, and Two-Way ANOVA Results

KR Condition	Experimental Condition: Self-Paced N	Mean	SD	Computer-Paced N	Mean	SD	Combined Conditions N	Mean	SD
KR	42	.54	.23	42	.53	.21	84	.54	.21
No-KR	34	.57	.24	35	.52	.24	69	.55	.24
Total Group	76	.56	.23	77	.53	.22	53	.54	.23

Two-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	2	.02	.31	.735
KR	1	.00	.06	.802
Pacing	1	.03	.56	.457
Two-Way Interaction				
KR × Pacing	1	.02	.31	.581
Residual	149	.05		
Total	152	.05		

[a]Probability of error in rejecting null hypothesis.

Response Pattern Information

Means and standard deviations of response pattern information as a function of testing strategy, KR, and pacing conditions are shown in Table 4. The results of the three-way analysis of variance are also shown. As Table 4 indicates, there was a significant main effect for testing strategy and a significant Test · KR · Pacing interaction, which is plotted in Figure 1.

The main effect for testing strategy indicated that mean response pattern information was highest (8.67) in the fixed-length stradaptive testing condition, next highest (6.44) in the variable-length stradaptive testing condition, and lowest (4.20) in the conventional testing condition. Post hoc analysis indicated that mean level of observed information of the conventional test was significantly less ($p \leq .01$) than either of the stradaptive tests and that the fixed-length stradaptive test was significantly higher ($p \leq .01$) than the variable-length stradaptive test. These data indicate that the conventional test would have to be 103 items long in order to obtain the same level of information/precision as did the 50-item fixed-length stradaptive test, or 77 items long to measure with the same degree of precision/information as did the variable-length stradaptive test that had a mean test length of approximately 27 items (SD = 4.4).

The three-way interaction data in Figure 1 show different effects on mean information as a function of KR and pacing conditions. KR and pacing conditions

Table 3
Means and Standard Deviations of Mean Difficulty Correct Scores for Stradaptive Tests With and Without KR in Computer- and Self-Paced Conditions, and Three-Way ANOVA Results

Test and KR Condition	Experimental Condition: Self-Paced N	Mean	SD	Computer-Paced N	Mean	SD	Combined Conditions N	Mean	SD
Stradaptive: Fixed Length									
KR	41	-.14	1.15	41	-.22	.96	82	-.18	1.06
No-KR	34	-.51	1.01	33	.15	1.02	67	-.18	1.07
Stradaptive: Variable Length									
KR	39	-.09	.87	40	-.10	1.09	79	-.10	.98
No-KR	33	-.21	1.13	33	-.15	1.09	66	-.18	1.10
Combined Groups									
Stradaptive									
Fixed Length	75	-.31	1.10	74	-.05	1.00	149	-.18	1.06
Variable Length	72	-.14	.99	73	-.12	1.08	145	-.13	1.04
KR	80	-.12	1.02	81	.16	1.02	161	-.14	1.02
No-KR	67	-.36	1.08	66	.01	1.06	133	-.18	1.08
Total Group	147	-.23	1.05	147	-.09	1.04	294	-.16	1.05

Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	3	.58	.54	.658
Test	1	.18	.16	.689
KR	1	.11	.10	.751
Pacing	1	1.46	1.34	.249
Two-Way Interactions	3	1.46	1.34	.263
Test × KR	1	.14	.13	.719
Test × Pacing	1	1.07	.98	.323
KR × Pacing	1	3.20	2.93	.088
Three-Way Interaction				
Test × KR × Pacing	2	2.09	1.91	.167
Residual	286	1.09		
Total	293	1.09		

[a] Probability of error in rejecting null hypothesis.

had no effect on the information for the conventional test. For the variable-length stradaptive test, slight differences in information levels were observed for the KR conditions, but pacing conditions had no differential effects. However, for the fixed-length stradaptive test, pacing and KR conditions interacted with respect to information. Highest mean information values were observed for the computer-paced no-KR condition, and lowest mean information was observed for the computer-paced KR condition; mean information values for the self-paced condition were intermediate between those for the computer-paced condition, but in the self-paced condition KR had opposite effects.

Table 4
Means and Standard Deviations of Response Pattern Information for Conventional and Stradaptive Tests With and Without KR in Computer- and Self-Paced Conditions, and Three-Way ANOVA Results

Test and KR Condition	Experimental Condition: Self-Paced N	Self-Paced Mean	Self-Paced SD	Experimental Condition: Computer-Paced N	Computer-Paced Mean	Computer-Paced SD	Combined Conditions N	Combined Conditions Mean	Combined Conditions SD
Conventional									
KR	35	4.21	2.19	36	4.48	1.92	71	4.35	2.04
No-KR	28	4.01	2.41	28	4.03	2.38	56	4.02	2.37
Stradaptive: Fixed Length									
KR	40	9.18	2.92	38	7.18	2.93	78	8.21	3.07
No-KR	33	8.22	2.90	33	10.20	4.71	66	9.21	4.01
Stradaptive: Variable Length									
KR	38	5.82	3.57	40	5.65	4.25	78	5.73	3.91
No-KR	31	6.70	4.54	32	6.18	5.23	63	6.44	4.87
Combined Groups									
Conventional	63	4.11	2.25	64	4.25	2.15	127	4.20	2.19
Stradaptive									
Fixed Length	73	8.20	2.91	63	8.69	3.82	144	8.67	3.56
Variable Length	69	6.26	4.05	72	5.91	4.74	141	6.44	4.87
KR	113	6.51	3.60	114	5.79	3.37	227	6.15	3.50
No-KR	92	6.43	3.80	93	6.96	5.02	185	6.70	4.45
Total Group	205	6.47	3.70	207	6.38	4.19	412	6.43	3.98

Three-Way Analysis of Variance

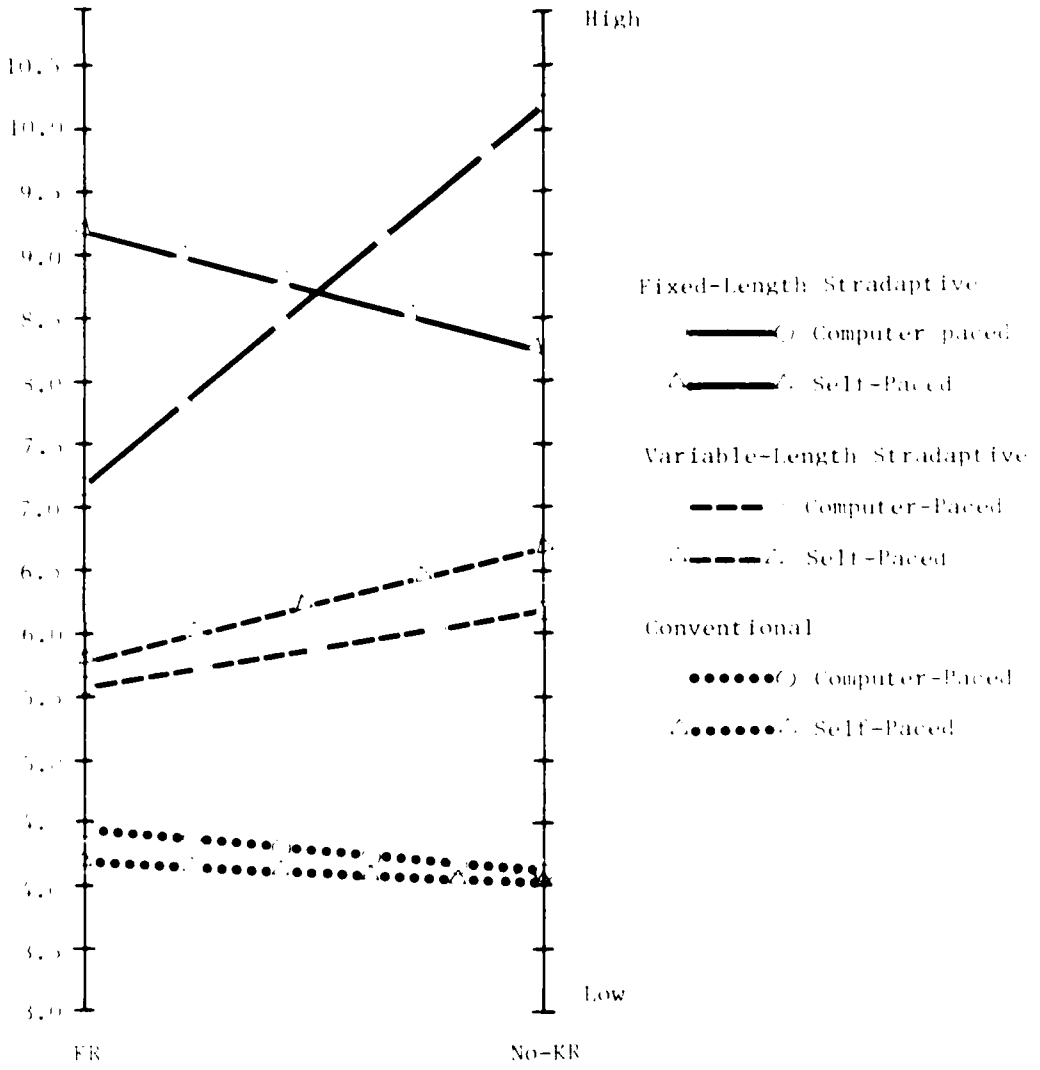
Source of Variation	df	Mean Square	F	p[a]
Main Effects	4	349.09	28.67	.001
Test	2	681.75	55.99	.001
KR	1	24.71	2.03	.155
Pacing	1	1.47	.12	.728
Two-Way Interactions	5	14.53	1.19	.311
Test × KR	2	16.06	1.32	.269
Test × Pacing	2	2.12	.18	.840
KR × Pacing	1	36.68	3.01	.083
Three-Way Interaction				
Test × KR × Pacing	2	53.26	4.37	.013
Residual	400	12.18		
Total	411	15.68		

[a]Probability of error in rejecting null hypothesis.

Response Latency

Means and standard deviations for the average latency over all items as a function of testing strategy, KR, and pacing conditions are presented in Table 5. The results of the three-way analysis of variance are also shown. As can be

Figure 1
Mean Response Pattern Information as a Function of
Testing Strategy, KR, and Pacing Conditions



seen in Table 5, there was a significant main effect on mean latency for testing strategy. No other main effects or interactions were significant.

Average time over all items for completion of the items was largest (15.79 sec.) in the variable-length stradaptive test, smallest (14.22 sec.) in the conventional test condition, and intermediate (14.75 sec.) in the fixed-length stradaptive testing condition. Post hoc analysis indicated that there was a significant difference ($p \leq .01$) in average latency between the variable-length stradaptive and conventional testing conditions.

Table 5
Means and Standard Deviations Over All Items of Average Response Latencies for Conventional and Stradaptive Tests With and Without KR in Computer- and Self-Paced Conditions, and Three-Way ANOVA Results

	Experimental Condition						Combined Conditions		
	Self-Paced			Computer-Paced					
Test and KR Condition	N	Mean	SD	N	Mean	SD	N	Mean	SD
Conventional									
KR	42	14.18	5.66	42	14.56	5.05	84	14.37	5.33
No-KR	34	13.97	5.67	35	14.11	4.08	69	14.04	4.89
Stradaptive: Fixed Length									
KR	41	13.42	3.72	41	14.74	5.44	82	14.08	4.68
No-KR	34	15.43	5.78	33	15.73	5.14	67	15.58	5.44
Stradaptive: Variable Length									
KR	39	15.22	4.52	40	15.50	4.63	79	15.36	4.55
No-KR	33	16.73	6.17	33	15.86	5.72	66	16.30	5.92
Combined Groups									
Conventional	76	14.09	5.63	77	14.36	4.61	153	14.22	5.12
Stradaptive									
Fixed Length	75	14.33	4.84	74	15.19	5.29	149	14.75	5.07
Variable Length	72	15.91	5.36	73	15.66	5.12	145	15.79	5.22
KR	122	14.25	4.73	123	14.93	5.03	245	14.59	4.88
No-KR	101	15.36	5.93	101	15.21	5.02	202	15.29	5.48
Total Group	233	14.76	5.32	224	15.06	5.02	447	14.91	5.17

Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	4	62.52	2.36	.053
Test	2	93.07	3.51	.031
KR	1	53.05	2.00	.158
Pacing	1	10.09	.38	.538
Two-Way Interactions	5	21.27	.80	.549
Test · KR	2	32.89	1.24	.291
Test · Pacing	2	11.22	.42	.655
KR · Pacing	1	17.70	.67	.415
Three-Way Interaction				
Test · KR · Pacing	2	2.53	.08	.919
Residual	435	26.55		
Total	446	26.70		

[a]Probability of error in rejecting null hypothesis.

Psychological Reactions

Motivation. The means, standard deviations, and three-way analysis of variance are presented in Table 6 for reported motivation level as a function of testing strategy, KR, and pacing conditions. There was no main effect on re-

ported motivation level due to testing strategy or KR condition. There was, however, a significant KR · Pacing interaction, which is plotted in Figure 2. The figure shows that reported motivation was high under computer-paced conditions when KR was given, but low under no-KR conditions. In the self-paced condition, however, the opposite relationship was found. When tests were self-paced, motivation was lower under KR than under no-KR conditions.

Table 6
Means and Standard Deviations of Motivation Scores for Conventional and Stradaptive Tests With and Without KR in Computer- and Self-Paced Conditions, and Three-way ANOVA Results

	Experimental Condition						Combined Conditions		
	Self-Paced			Computer-Paced					
Test and KR Condition	N	Mean	SD	N	Mean	SD	N	Mean	SD
Conventional									
KR	42	.08	2.04	42	.12	2.44	84	.10	2.23
No-KR	34	-.09	2.18	35	-.46	2.36	69	-.28	2.26
Stradaptive: Fixed Length									
KR	41	-.52	2.20	41	.24	2.28	82	-.14	2.26
No-KR	34	.17	2.05	33	-.30	2.12	67	-.06	2.08
Stradaptive: Variable Length									
KR	39	.22	1.66	40	-.01	2.05	79	.11	1.86
No-KR	33	.19	1.55	33	-.78	2.11	66	-.29	1.90
Combined Groups									
Conventional	76	.01	2.09	77	-.14	2.41	153	-.07	2.25
Stradaptive									
Fixed Length	75	-.21	2.15	74	.00	2.21	149	-.10	2.18
Variable Length	72	.21	1.60	73	-.36	2.09	145	-.08	1.88
KR	122	-.08	1.99	123	.12	2.25	245	.02	2.12
No-KR	101	.09	1.93	101	-.51	2.19	202	-.21	2.08
Total Group	223	-.00	1.96	224	-.16	2.24	447	-.08	2.11

Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	4	2.29	.52	.725
Test	2	.06	.01	.987
KR	1	6.04	1.36	.244
Pacing	1	3.02	.68	.410
Two-Way Interactions	5	6.72	1.51	.184
Test · KR	2	2.58	.58	.560
Test · Pacing	2	5.50	1.24	.291
KR · Pacing	1	17.29	3.89	.049
Three-Way Interaction				
Test · KR · Pacing	2	1.61	.36	.696
Residual	435	4.44		
Total	446	4.43		

[a]Probability of error in rejecting null hypothesis.

Figure 2
Mean Motivation Scores as a Function of
KR and Pacing Conditions

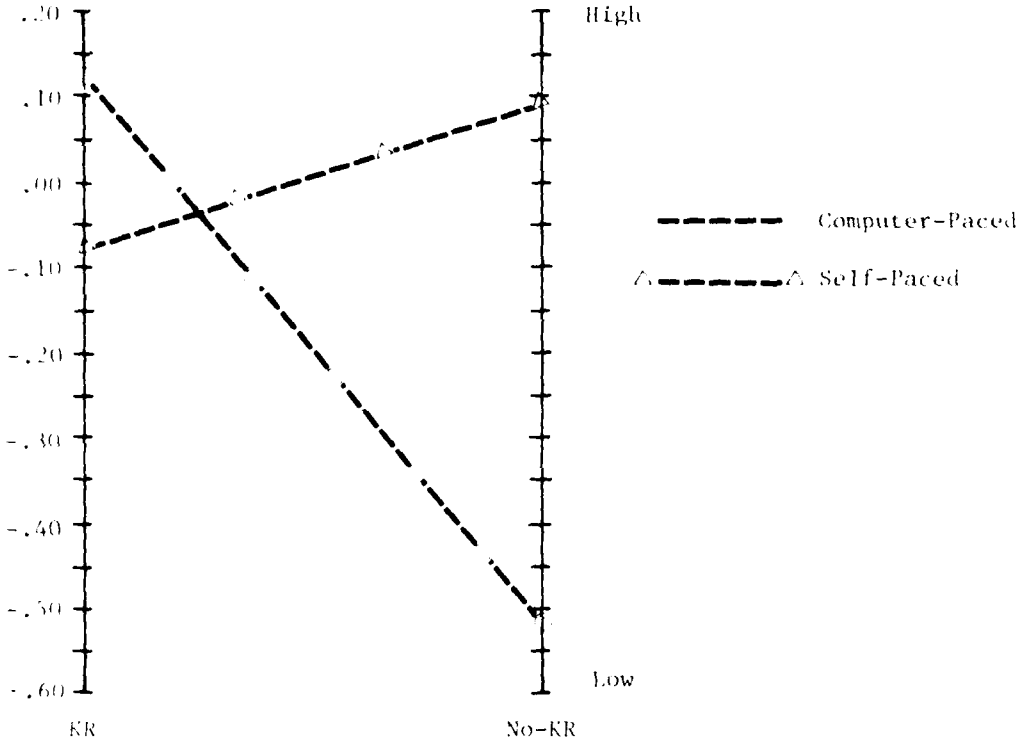


Figure 2 also indicates that there was little difference in motivation between computer- and self-paced conditions when KR was given. In post hoc analysis of the mean motivation scores, the sum of squares of the two pacing means under KR conditions was compared in an F ratio to the residual sum of squares with 1 and 435 df. In a similar manner, the sum of squares of the two pacing means under no-KR conditions were compared to the residual sum of squares. This analysis indicated that there was a significant difference among no-KR means ($F = 4.094$, $p \leq .05$) but not among KR means ($F = .5518$). Thus, when students received feedback on test performance, the average levels of motivation they reported were relatively high and were similar regardless of the pacing of item presentation. Under no-KR conditions, however, motivation varied greatly and significantly as a function of pacing condition.

Table 7 shows the percentage of students selecting each alternative of the Motivation Scale items in each KR condition, testing strategy (conventional vs. stradaptive), and pacing condition and in the total sample, and the results of chi-square tests of independence within experimental conditions. In general, students reported a relatively high level of motivation as assessed by the three items that defined the Motivation Scale. Approximately 60% indicated on Question 6 that they "almost always" were careful to select the best alternative to a question. When asked if they were challenged to do well on the test (Question 13), nearly three-quarters of the students replied that they were "fairly much"

Table 7
Response Percentages for Motivation Questions as a Function of KR Condition, Testing Strategy, and Pacing Condition, and for Total Group

Question	KR Condition			Testing Strategy			Pacing Condition			Total Group
	KR	No KR	p[a]	Conventional	Stradaptive	p[a]	Self-Paced	Computer-Paced	p[a]	
6. How frequently were you careful to select what you thought was the best answer to each question?										
1. Almost always	62.4	61.4	.95	63.4	61.2	.59	63.2	60.7	.87	62.0
2. Frequently	29.0	29.2		26.8	30.3		27.8	30.4		29.1
3. Sometimes	7.8	8.9		8.5	8.2		8.1	8.5		8.3
4. Rarely	.8	.5		1.4	.3		.9	.4		.7
5. Never	0	0		0	0		0	0		0
13. Did you feel challenged to do as well as you could on the test?										
1. Not at all	3.7	5.0	.68	4.6	4.1	.99	2.7	5.8	.31	4.3
2. Somewhat	26.1	29.2		27.5	27.6		27.8	27.2		27.5
3. Fairly much so	39.2	39.1		38.6	39.5		41.7	36.6		39.1
4. Very much so	31.0	26.7		29.4	28.9		27.8	30.4		29.1
18. Did you care how well you did on the test?										
1. I cared a lot	23.7	18.8	.21	22.2	21.1	.99	22.9	20.1	.51	21.5
2. I cared some	51.8	51.0		51.0	51.7		53.4	49.6		51.5
3. I cared a litte	113.5	21.3		17.0	17.0		13.9	20.1		17.0
4. I cared very little	8.6	6.4		7.8	7.5		7.6	7.6		7.6
5. I didn't care at all	2.4	2.5		2.0	2.7		2.2	2.7		2.5

[a]Probability of error in rejecting null hypothesis of independence, based on chi-square analysis.

or "very much" challenged. About 70% of the sample indicated that they cared "some" or "a lot" about how well they did on the experimental test (Question 18). There were no statistically significant differences between KR, testing, or pacing conditions on any of the Motivation Scale items.

Anxiety. Means and standard deviations of reported anxiety level as a function of each test, KR, and pacing combination are presented in Table 8. Also shown is the three-way analysis of variance. The analysis indicates that mean anxiety scores yielded a significant three-way (Test × KR × Pacing) interaction. A diagram of the interaction is in Figure 3. In comparison with those in the KR condition, students not receiving KR reported higher anxiety in taking computer-paced variable-length stradaptive and conventional tests and in both self-paced stradaptive tests. Lower levels of anxiety in no-KR conditions were reported in the computer-paced conventional testing conditions. Students receiving KR, however, reported about the same level of anxiety regardless of testing conditions.

In post hoc analysis of the mean anxiety scores, the sum of squares of the six Test × Pacing means under KR conditions was compared in an F ratio to the residual sum of squares with 5 and 435 df, and the sum of squares of the six Test × Pacing means under no-KR conditions were compared to the residual sum of squares. This analysis showed that the differences among the six KR means were not statistically significant ($F = .339$), whereas the differences among no-KR means were statistically significant ($F = 2.96$, $p \leq .05$). A difference between any pair of mean anxiety scores of 1.52 or greater was statistically significant in the no-KR condition. These data show, therefore, that mean anxiety scores did not differ significantly in the no-KR condition as a function of pacing conditions for either of the stradaptive tests but that a significant difference did occur as a result of pacing (in the no-KR condition) for the conventional test. Thus, there was no significant variation in mean anxiety scores among testing conditions when students received KR; but when KR was not provided, levels of anxiety varied significantly as a function of testing condition, with significant differences occurring only for the conventional test as a function of pacing conditions.

Table 9 shows the percentage of students in each experimental condition and in the total group who chose each item alternative to the four anxiety items, and the results of chi-square tests within experimental conditions. Overall, students reported a low level of anxiety. Approximately 68% of the total sample reported on Question 4 that they did not worry "at all" or worried "somewhat" during testing. When asked if they were nervous while taking the test (Question 7), about 60% answered that they were not nervous at all. Most students (45%) indicated on Question 11 that they were relaxed during testing, but some (36%) reported that they were neither tense nor relaxed. Approximately 92% of the total group expressed doubt that nervousness prevented them from doing well on the test (Question 16). The only statistically significant difference was observed on Question 11--between the KR conditions. Students in the KR group tended to report lower levels of being "tense" or "very tense" than did those in the no-KR group.

Difficulty perception. Means and standard deviations for difficulty perception scores as a function of test, KR, and pacing conditions are presented in Table 10; also shown are the results from the three-way analysis of variance.

Table 8
Means and Standard Deviations of Anxiety Scores for Conventional and Stradaptive Tests With and Without KR in Computer- and Self-Paced Conditions, and Three-way ANOVA Results

Test and KR Condition	Experimental Condition: Self-Paced N	Mean	SD	Experimental Condition: Computer-Paced N	Mean	SD	Combined Conditions N	Mean	SD
Conventional									
KR	42	-3.15	2.69	42	-3.38	2.76	84	-3.26	2.71
No-KR	34	-4.49	2.43	35	-2.46	3.23	69	-3.46	3.02
Stradaptive: Fixed Length									
KR	41	-3.62	2.98	41	-2.99	3.10	82	-3.31	3.03
No-KR	34	-3.28	2.55	33	-4.19	2.72	67	-3.73	2.66
Stradaptive: Variable Length									
KR	39	-3.67	3.07	40	-3.26	2.50	79	-3.46	2.79
No-KR	33	-2.47	3.40	33	-2.98	3.32	66	-2.73	3.35
Combined Groups									
Conventional	76	-3.75	2.64	77	-2.96	3.00	153	-3.35	2.85
Stradaptive									
Fixed Length	75	-3.47	2.78	74	-3.53	2.98	149	-3.50	2.87
Variable Length	72	-3.12	3.26	73	-3.14	2.88	145	-3.13	3.07
KR	122	-3.48	2.90	123	-3.21	2.78	245	-3.34	2.84
No-KR	101	-3.42	2.91	101	-3.20	3.16	202	-3.31	3.03
Total Group	223	-3.45	2.90	224	-3.21	2.95	447	-3.33	2.92

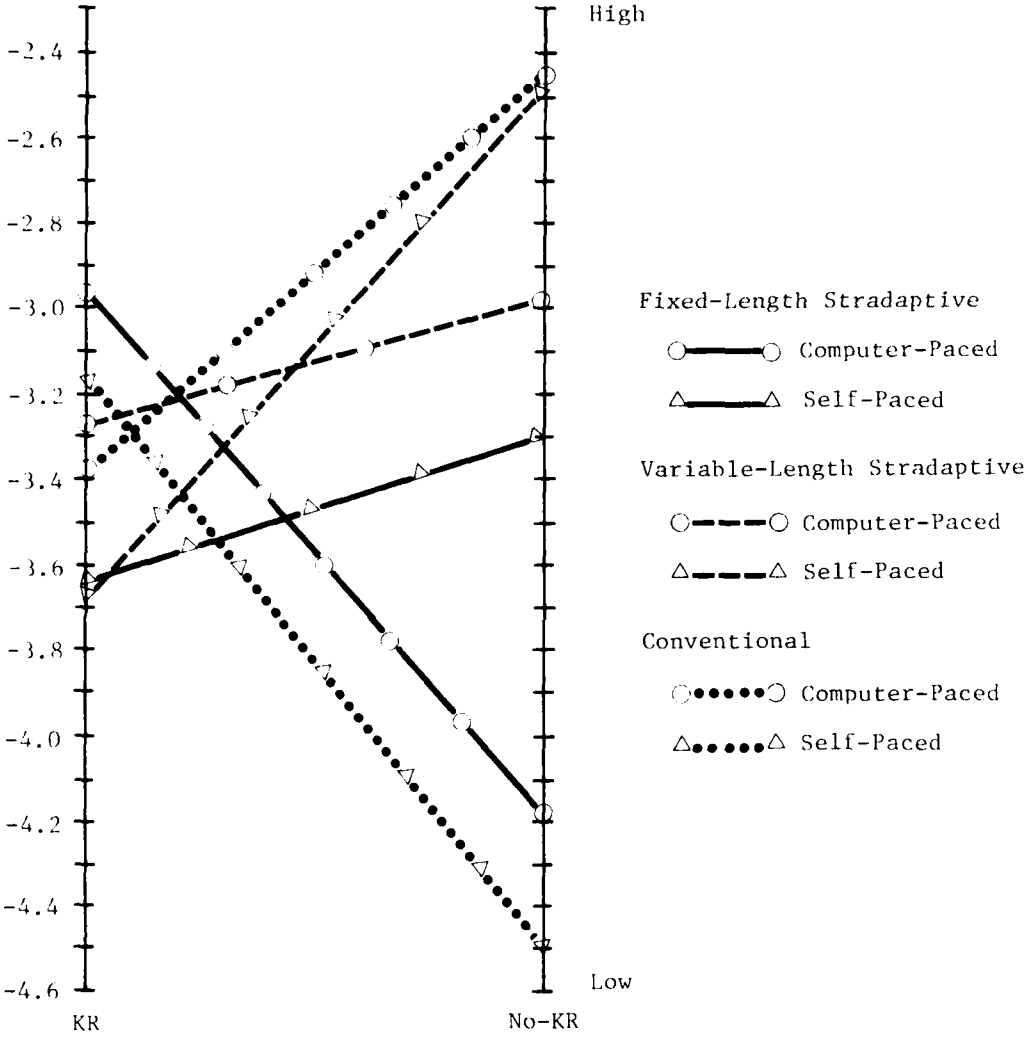
Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	4	4.25	.50	.733
Test	2	5.05	.60	.550
KR	1	.12	.01	.906
Pacing	1	6.69	.79	.374
Two-Way Interactions	5	9.04	1.07	.376
Test × KR	2	13.82	1.64	.196
Test × Pacing	2	8.79	1.04	.354
KR × Pacing	1	.05	.01	.940
Three-Way Interaction				
Test × KR × Pacing	2	38.86	4.60	.011
Residual	435	8.44		
Total	446	8.55		

[a]Probability of error in rejecting null hypothesis.

There were no significant main effects for test, KR, or pacing conditions on perception of test difficulty. The Test × KR interaction, however, approached significance ($p \leq .086$). The Test × KR interaction indicated that when KR was provided, students taking the conventional test perceived it to be less difficult (mean = .26) than students not receiving KR (mean = 1.79). Essentially

Figure 3
Mean Anxiety Scores as a Function of
Testing Strategy, KR, and Pacing Conditions



equal levels of difficulty perception were reported by stradaptive testees under KR and no-KR conditions, with mean levels of difficulty perception of .86 and 1.10, respectively.

Table 11 presents the percentages of the experimental groups and of the total group selecting each response alternative on the six Difficulty Perception Scale items, and the results of the chi-square tests within experimental conditions. In general, most students felt that the test items were seldom easy and frequently too hard (Question 1) and that the test was too difficult in relation to their vocabulary ability (Question 2). Most (54%) felt "somewhat" frustrated by the difficulty of the test questions (Question 12). The chi-square analyses show that on every Difficulty Perception Scale item, there was a significant

Table 9

Response Percentages for Anxiety Questions as a Function of KR Condition, Testing Strategy, and Pacing Condition, and for Total Group

Question	KR Condition			Testing Strategy			Pacing Condition			Total Group
	KR	No KR	p[a]	Conventional	Stradaptive	p[a]	Self-Paced	Computer-Paced	p[a]	
4. During testing did you worry about how well you you would do?										
1. Not at all	15.9	16.3	.88	15.0	16.7	.92	15.7	16.5	.36	16.1
2. Somewhat	53.9	51.0		51.6	53.1		54.3	50.9		52.6
3. Fairly much	21.2	24.3		24.2	21.8		19.7	25.4		22.6
4. Very much	9.0	8.4		9.2	8.5		10.3	7.1		8.7
7. Were you nervous while taking the test?										
1. Not at all	59.2	58.4	.76	61.4	57.5	.54	61.4	56.3	.38	58.8
2. Somewhat	32.7	33.2		28.8	35.0		30.9	34.8		32.9
3. Moderately so	6.9	5.9		7.8	5.8		6.7	6.3		6.5
4. Very much so	1.2	2.5		2.0	1.7		.9	2.7		1.8
11. How did you feel while taking the test?										
1. Very tense	.8	2.0	.05	1.3	1.4	.99	1.3	1.3	.50	1.3
2. Somewhat tense	15.9	20.8		18.3	18.0		16.1	20.1		18.1
3. Neither tense nor relaxed	36.7	34.2		35.9	35.4		34.5	36.6		35.6
4. Somewhat relaxed	31.8	21.8		26.1	27.9		30.9	23.7		27.3
5. Very relaxed	14.7	21.3		18.3	17.3		17.0	18.3		17.7
16. Did nervousness while taking the test prevent you from doing your best?										
1. Yes, definitely	.4	1.0	.53	0	1.0	.45	1.3	0	.17	.7
2. Yes, somewhat	6.9	6.9		7.2	6.8		5.8	8.0		6.9
3. Probably not	43.7	37.6		37.9	42.5		38.6	43.3		40.9
4. Definitely not	49.0	54.5		54.9	49.7		54.3	48.7		51.5

[a]Probability of error in rejecting null hypothesis of independence, based on chi-square analysis.

Table 10
Means and Standard Deviations of Difficulty Perception Scores for Conventional and Stradaptive Tests With and Without KR In Computer- and Self-Paced Conditions, and Three-Way ANOVA Results

Test and KR Condition	Experimental Condition: Self-Paced N	Mean	SD	Computer-Paced N	Mean	SD	Combined Conditions N	Mean	SD
Conventional									
KR	42	.04	4.66	42	.47	4.56	84	.26	4.58
No-KR	34	.88	4.99	35	2.69	4.75	69	1.79	4.92
Stradaptive: Fixed Length									
KR	41	.69	2.89	41	1.32	2.78	82	1.01	2.83
No-KR	34	.62	3.21	33	1.11	3.53	67	.86	3.36
Stradaptive: Variable Length									
KR	39	1.40	3.23	40	.81	3.17	79	1.10	3.19
No-KR	33	.44	3.42	33	1.40	3.45	66	.92	3.44
Combined Groups									
Conventional	76	.41	4.79	77	1.48	4.75	153	.95	4.78
Stradaptive									
Fixed Length	75	.66	3.02	74	1.22	3.12	149	.94	3.07
Variable Length	72	.96	3.33	73	1.07	3.29	145	1.02	3.31
KR	122	.69	3.70	123	.86	3.58	245	.78	3.64
No-KR	101	.65	3.92	101	1.75	3.99	202	1.20	3.99
Total Group	223	.67	3.80	224	1.26	3.79	447	.97	3.80

Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	4	14.74	1.02	.394
Test	2	.23	.02	.984
KR	1	19.67	1.37	.243
Pacing	1	38.89	2.70	.101
Two-Way Interactions	5	22.30	1.55	.173
Test · KR	2	35.39	2.46	.086
Test · Pacing	2	8.31	.58	.561
KR · Pacing	1	23.73	1.65	.200
Three-Way Interaction				
Test · KR · Pacing	2	7.91	.55	.577
Residual	435	14.38		
Total	446	14.44		

[a]Probability of error in rejecting null hypothesis.

difference between conventional and stradaptive testing conditions. Inspection of the distribution of percentages indicates that students in the stradaptive testing conditions perceived their test as being of more appropriate difficulty than students taking the conventional test. The conventional test was more often perceived as being either too easy or too hard by the examinees. There

Table 11

Response Percentages for Difficulty Perception Questions as a Function of KR Condition, Testing Strategy, and Pacing Condition, and for Total Group

Question	KR Condition			Testing Strategy			Pacing Condition			Total Group
	KR	No KR	p[a]	Conventional	Stradaptive	p[a]	Self-Paced	Computer-Paced	p[a]	
1. How often did you feel that the questions in the test were too easy for you?										
1. Always	0	0	.70	0	0	.02	0	0	.22	0
2. Frequently	4.9	3.5		7.8	2.4		4.9	3.6		4.3
3. Sometimes	33.1	37.1		30.7	37.1		39.0	30.8		34.9
4. Seldom	50.6	47.0		47.1	50.0		44.8	53.1		49.0
5. Never	11.4	12.4		14.4	10.5		11.2	12.5		11.9
2. How often did you feel that the questions in the test were too hard for you?										
1. Always	.8	1.5	.07	2.6	.3	<.01	.9	1.3	.53	1.1
2. Frequently	48.2	52.5		49.7	50.3		52.5	47.8		50.1
3. Sometimes	39.2	37.1		29.4	42.9		35.0	41.5		38.3
4. Seldom	11.8	6.9		16.3	6.1		10.3	8.9		9.6
5. Never	0	2.0		2.0	.3		1.3	.4		.9
5. On how many questions did you guess?										
1. Almost all of the questions	1.2	5.5	.10	6.0	1.7	<.01	2.3	4.0	.58	2.3
2. More than half of the questions	15.6	13.5		19.2	12.3		12.7	16.6		12.7
3. About half of the questions	30.5	26.0		23.2	31.2		29.1	27.8		29.1
4. Less than half of the questions	35.0	34.5		25.8	39.4		35.5	34.1		35.5
5. Almost none of the questions	17.7	20.5		25.8	15.4		20.5	17.5		20.5
6. None of the questions	0	0		0	0		0	0		0

- continued on the next page -

Table 11, continued
Response Percentages for Difficulty Perception Questions as a Function of KR Condition, Testing Strategy, and Pacing Conditon, and for Total Group

Question	KR Condition			Testing Stategy			Pacing Condition			Total Group
	KR	No KR	p[1]	Conven- tional	Strad- aptive	p[1]	Self- Paced	Computer- Paced	p[1]	
8. How often were you sure that your answers to the questions were correct?										
1. Almost always	6.9	5.9	.11	11.8	3.7	<.01	9.0	4.0	.16	9.0
2. More than half of the time	21.6	19.3		22.9	19.4		20.2	21.0		20.2
3. About half of the time	35.5	38.1		22.2	44.2		38.6	34.8		38.6
4. Less than half of the time	32.7	27.7		34.0	28.6		26.9	33.8		26.9
5. Almost Never	3.3	8.9		9.2	4.1		5.4	6.3		5.4
10. In relation to your vocabulary ability, how difficult was the test for you?										
1. Much too difficult	7.8	9.9	.65	15.7	5.1	<.01	7.6	9.8	.28	8.7
2. Somewhat too difficult	52.7	55.9		48.4	57.1		50.2	58.0		54.1
3. Just about right	35.1	31.7		29.4	35.7		38.1	29.0		33.6
4. Somewhat too easy	3.7	2.0		4.6	2.0		3.1	2.7		2.9
5. Much too easy	.8	.5		2.0	0		.9	.4		.7
12. Did you feel frustrated by the difficulty of the test questions?										
1. Not at all	35.1	28.2	.41	37.9	28.9	.01	31.4	32.6	.76	32.0
2. Somewhat	51.4	55.9		44.8	58.5		54.7	52.2		53.5
3. Fairly much so	10.6	11.4		12.4	10.2		11.2	10.7		11.0
4. Very much so	2.9	4.5		5.9	2.4		2.7	4.5		3.6

[1]Probability of error in rejecting null hypothesis of independence based on chi-square analysis.

were no significant differences in item response distributions within the KR or pacing conditions.

KR Reaction. Table 12 shows the means, standard deviations, and two-way analysis of variance for KR reaction as a function of pacing and test strategy conditions. The table indicates no significant main effects or interactions among these variables for KR reaction.

Table 12
Means and Standard Deviations of KR Reaction Scores for Conventional and Stradaptive Tests in Computer- and Self-Paced Conditions, and Two-Way ANOVA Results

	Experimental Condition						Combined Conditions		
	Self-Paced			Computer-Paced					
Test	N	Mean	SD	N	Mean	SD	N	Mean	SD
Conventional	42	-.73	1.59	42	-.45	1.03	84	-.59	1.34
Stradaptive									
Fixed Length	41	-.81	1.70	41	-.63	.96	82	-.72	1.38
Variable Length	39	-.46	.98	40	-.48	1.42	79	-.52	1.22
Total Group	122	-.67	1.46	123	-.55	1.14	245	-.61	1.31

Three-Way Analysis of Variance

Source of Variation	df	Mean Square	F	p[a]
Main Effects	3	.85	.49	.689
Test	2	.83	.48	.621
Pacing	1	.89	.51	.474
Two-Way Interaction				
Test · Pacing	2	.85	.49	.614
Residual	239	1.73		
Total	244	1.72		

[a]Probability of error in rejecting null hypothesis.

Items assessing the reactions to feedback were administered to students in the KR condition. Table 13 gives the percentage of students in the KR condition selecting each alternative of the five KR Reaction Scale questions. Overall, the reaction to feedback was very favorable. Approximately 80% of the KR-condition students indicated that feedback made testing much more interesting (Question 19) and that they were interested in knowing whether their answers were right or wrong (Question 24). About 80% of the students indicated that feedback did not interfere with their ability to concentrate on the test nor make them nervous (Question 20). About 93% of the KR-condition students said they liked getting the feedback (Question 26). Chi-square analysis of the KR Reaction Scale questions indicated that of the students receiving KR, students in the computer-paced condition were more often "very interested" in knowing whether their answers were right or wrong (Question 24). There were no other significant differences in response distributions within the experimental conditions.

Table 13
Response Percentages of KR Reaction Items as a Function of Test and Pacing Conditions, and for Total Group

Question	Testing Strategy: Conventional	Testing Strategy: Stradaptive	p[a]	Pacing Condition: Self-Paced	Pacing Condition: Computer-Paced	p[a]	Total Group: N	Total Group: %
19. Did getting feedback on this test make it more interesting or less interesting?								
1. Much more interesting	78.8	80.0	.38	78.9	80.3	.60	121	79.6
2. Somewhat more interesting	19.2	20.0		19.7	19.7		30	19.7
3. Didn't make any difference	0.0	0.0		0.0	0.0		0	0.0
4. Somewhat less interesting	1.9	0.0		1.3	0.0		1	.7
5. Much less interesting	0.0	0.0		0.0	0.0		0	0.0
20. Did receiving feedback after each question interfere with your ability to concentrate on the test?								
1. No, not at all	88.5	83.0	.32	84.2	85.5	.21	129	84.9
2. Yes, somewhat	7.7	15.0		14.5	10.5		19	12.5
3. Yes, moderately so	3.8	1.0		0.0	3.9		3	2.0
4. Yes, very much so	0.0	1.0		1.3	0.0		1	.7
21. Did getting feedback after each question make you nervous?								
1. No, not at all	73.1	76.0	.29	76.3	73.7	.71	114	75.0
2. Yes, somewhat	25.0	21.0		22.4	22.4		34	22.4
3. Yes, moderately so	0.0	3.0		1.3	2.6		3	2.0
4. Yes, very much so	1.9	0.0		0.0	1.3		1	.7
24. Were you interested in knowing whether your answers were right or wrong?								
1. I was very interested	90.4	79.0	.08	76.3	89.5	.04	126	82.9
2. I was moderately interested	5.8	19.0		18.4	10.5		22	14.5
3. I was somewhat interested	3.8	2.0		5.3	0.0		4	2.6
4. I didn't care at all	0.0	0.0		0.0	0.0		0	0.0
26. How did you feel about getting feedback?								
1. I'd rather not know whether my answers were right or wrong	5.8	3.0	.25	3.9	3.9	.60	6	3.9
2. I really don't care whether I get feedback or not	0.0	4.0		3.9	1.3		4	2.6
3. I liked getting the feedback	94.2	93.0		92.1	94.7		142	93.4

[a]Probability of error in rejecting null hypotheses of independence, based on chi-square analysis.

Other psychological reactions questions. The eight questions that were not included in the other four psychological reactions scales are shown in Table 14. Also shown are the percentage of students in the KR, testing strategy, and pacing conditions as well as in the total group who selected each alternative of the items, and the results of the chi-square tests within experimental conditions. The chi-square results for the non-KR items were based on data from 446 students, and the chi-square analyses for the KR items were based on data from 152 students.

Three items showed significant differences for the KR conditions, one was significant for testing strategies, but no significant chi-squares were observed between pacing conditions. More students (25.5%) taking the conventional test thought that the difficulty of the test was "seldom" or "never" right for someone of their ability (Question 3), but only 12.9% of the students taking the stradaptive test responded in these categories. More students (20%) receiving KR felt that they could have done better on the test if they had tried harder (Question 9) than students not receiving KR (10.9%). When students receiving KR responded to a question for which they didn't have an answer (Question 14), 90.2% said that they chose the most reasonable choice, whereas only 70.3% of students not receiving KR responded this way; more students in the no-KR condition (28.7%) answered such an item with the question mark key than did students who received KR (9.0%). Feedback also made students think that they did better on the test than students not receiving KR (Question 15).

Intercorrelations among Dependent Variables and Reliabilities

Table 15 shows the intercorrelations, levels of significance, and the numbers of students on which the correlations were based; internal consistency reliabilities are also shown for the four psychological reactions scales. Some variables were measured only under certain conditions, and for this reason correlations were based on differing numbers of subjects. Reaction to KR, for example, was obtained only from those in the KR condition. Since no students were administered both a stradaptive and a conventional test, there are no correlations between conventional test proportion-correct scores and stradaptive test mean difficulty correct scores.

In general, the results indicate that ability estimates correlated positively with scores on the Motivation Scale and negatively with the scores on the Anxiety and Difficulty Perception Scales. For example, the high-ability examinee reported higher levels of motivation, lower levels of anxiety, and perceived the test to be less difficult than the student of lower ability. Response pattern information, a measure of score precision, correlated positively with all ability estimates (r = .46 to .55), indicating that lower ability scores were less precise and that higher ability scores more precise. Reported motivation tended to correlate positively with ability estimates (r = .18 to .37), whereas reported anxiety had low positive correlations with ability estimates (r = -.11 to -.16). That students were able to accurately perceive the difficulty of the conventional test is reflected in the correlation of -.77 between proportion-correct scores and the scores on the Difficulty Perception Scale.

Correlations among the psychological reactions scales showed that persons who perceived the test to be difficult also tended to have high levels of reported anxiety (r = .25). Reported motivation had a slight negative correlation

Table 14

Response Percentages for Non-Scale Psychological Reactions Items as a Function of KR, Testing Strategy, and Pacing Condition, and for Total Group

Question	KR Condition			Testing Strategy			Pacing Condition			Total Group
	KR	No KR	p	Conventional	Stradaptive	p	Self-Paced	Computer-Paced	p	
3. How often did you feel that the questions in the test were just about right for someone of your ability?										
Always	3.3	1.5	.40	3.9	1.7	<.01	2.7	2.2	.34	2.5
Frequently	27.8	24.3		26.8	25.9		29.1	23.2		26.2
Sometimes	54.3	54.0		43.8	59.5		52.5	55.8		54.1
Seldom	13.5	18.3		22.2	12.2		13.5	17.9		15.6
Never	1.2	2.0		3.3	.7		2.2	.9		1.6
9. Do you think that you could have done better on the test if you tried harder?										
I definitely could have	4.5	3.5	.04	4.6	3.7	.34	3.1	4.9	.28	14.0
I probably could have	15.5	7.4		9.8	12.9		12.6	11.2		11.9
I'm not sure	27.3	23.8		29.4	23.8		25.6	25.9		25.7
I probably couldn't have	39.2	47.0		37.9	45.2		46.2	39.3		42.7
I definitely couldn't have	13.5	18.3		18.3	14.3		12.6	18.8		15.7
14. When you didn't know the answer to a question, what did you usually do?										
Picked one of the choices at random	.8	1.0	<.01	1.3	.7	.80	.4	1.3	.53	.9
Tried to pick the most reasonable choice	90.2	70.3		81.0	81.3		82.5	79.9		81.2
Answered using the question mark key	9.0	28.7		17.6	18.0		17.0	18.8		17.9
15. How well do you feel you did on this test in comparison to your performance on other tests like this?										
Much better	2.0	.5	.04	1.3	1.4	.39	.9	1.8	.84	1.3
Somewhat better	9.4	3.0		9.2	5.1		5.8	7.1		6.5
About the same	53.1	57.4		52.9	56.1		57.0	53.1		55.0
Somewhat worse	30.2	33.2		29.4	32.7		30.5	32.6		31.5
Much worse	5.3	5.9		7.2	4.8		5.8	5.4		5.6

- continued on next page -

Table 14, continued
Response Percentages for Non-Scale Psychological Reactions Items as a Function of KR, Testing Strategy, and Pacing Condition, and for Total Group

Question	KR Condition			Testing Strategy			Pacing Condition			Total Group
	KR	No KR	p	Conventional	Stradaptive	p	Self-Paced	Computer-Paced	p	
17. Did you feel that this test accurately measured your ability?										
It was very accurate	5.7	9.4	.11	8.5	6.8	.86	8.5	6.3	.15	7.4
It was somewhat accurate	37.6	33.7		37.3	35.0		38.6	33.0		35.8
It was somewhat inaccurate	27.8	20.3		21.6	25.9		23.8	25.0		24.4
It was very inaccurate	6.5	9.9		7.8	8.2		9.4	6.7		8.1
I really don't know	22.4	26.7		24.8	24.1		19.7	29.0		24.4
22. Did you try harder to get the questions right because you knew you would get feedback after each question?										
No, not at all				19.2	15.0	.13	14.5	18.4	.44	16.4
Yes, somewhat				26.9	33.0		36.8	25.0		30.9
Yes, moderately so				17.3	30.0		25.0	26.3		25.7
Yes, very much so				36.5	22.0		23.7	30.3		27.0
23. How often did you know whether your answer was right or wrong before you received the feedback?										
Almost always				11.5	8.0	.24	13.2	5.3	.37	9.2
Frequently				34.6	42.0		35.5	43.4		39.5
Sometimes				40.4	45.0		43.4	43.4		43.4
Almost never				13.5	5.0		7.9	7.9		7.9
25. How did you feel when you found that your answers were incorrect?										
It bothered me a lot				13.5	9.0	.22	9.2	11.8	.86	10.5
It bothered me some				48.1	42.0		42.1	46.1		44.1
It bothered me a little				25.0	41.0		38.2	32.9		35.5
It didn't bother me at all				13.5	8.0		10.5	9.2		9.9

Probability of error in rejecting null hypothesis of independence based on chi-square analysis.

Table 15

Intercorrelations of Dependent Variables (Lower Triangle)
Number of Observations on which Each Correlation is Based (Upper Triangle)
and Alpha Reliabilities for Psychological Reactions Scales (Main Diagonal)

Variable	1	2	3	4	5	6	7	8	9
Ability Estimates									
1. Maximum Likelihood Score		141	293	434	323	434	434	434	148
2. Proportion Correct Score	.90**			294	153	153	153	153	53
3. Mean Difficulty Correct	.97**			294	294	294	294	294	99
Other Ability Variables									
4. Average Response Latency	.01	-.25**	.07		447	447	447	447	151
5. Response Pattern Information	.46**	.56**	.55**	.04		447	447	447	151
Psychological Reactions Scales									
6. Difficulty Perception	-.38**	-.77**	-.11*	.20**	-.17**	.84	447	447	151
7. Anxiety	-.11**	-.16*	-.14**	.15**	-.07	.25**	.73	447	151
8. Motivation	.24**	.37**	.18**	.09*	.08*	-.13**	.15**	.61	151
9. KR Reactions	.03	.28*	-.07	-.18*	-.02	-.11	-.40**	.02	.61

*Significant at the .05 level.
**Significant at the .01 level.

with scores on the Difficulty Perception Scale. There was a slight but significant positive relationship ($r = .15$) between reported anxiety and motivation level. Reported anxiety also had a moderate negative correlation ($r = -.40$) with reaction to feedback. That is, the higher the level of anxiety, the less students liked getting feedback.

Table 15 also shows that the Difficulty Perception Scale, composed of six items, had an alpha coefficient of .84, whereas the four-item Anxiety Scale had a reliability of .73. Both the Motivation Scale, composed of three items and the five-item KR Reaction Scale had alpha reliabilities of .61.

DISCUSSION AND CONCLUSIONS

In prior research that investigated the use of feedback in adaptive testing (Betz & Weiss, 1976a, 1976b; Prestwood & Weiss, 1978), provision of feedback was confounded with rate of item presentation. The present research dealt with the unconfounded effects of KR and pacing as well as testing strategy on test performance and on psychological reactions to testing, including test anxiety and motivation.

Results that indicate the effect feedback had on test performance in adaptive testing have varied. The present research and a prior study (Prestwood & Weiss, 1978) did not replicate the finding that test performance was higher under KR than under no-KR conditions (Betz & Weiss, 1976a). The original study (Betz & Weiss, 1976a) differed from subsequent studies in that ability estimates were based on a combined group composed of low- and high-ability college students. The present research and the Prestwood and Weiss (1978) study, however, based ability estimates on groups composed only of high-ability college students. A significant increase in mean test performance under KR conditions has only been demonstrated when a relatively large range of college ability was tested.

The time a student spent solving each item was based on item latencies averaged across all items. These average item latencies were found to differ by type of testing strategy. Students took a significantly longer time (an average of 1.5 seconds per item) to solve items on the variable-length stradaptive test than they took on the conventional test. However, this difference may be due to variations in item difficulty on these tests. The mean difficulty of the conventional test was $\bar{b} = .02$, whereas the mean difficulties of the items administered in the fixed-length and variable-length stradaptive tests were $\bar{b} = .16$ and $\bar{b} = .26$, respectively. Thus, as might be expected, students taking the more difficult variable-length stradaptive test took longer, on the average, to respond to an item than those students taking the tests composed of easier items.

The longer latency of the stradaptive test was also found by Waters (1977) in comparison to a peaked conventional test but was not found by Betz (1976a). The difference in the findings was due to the difficulty of the conventional test in comparison to the stradaptive test for the particular ability of the groups tested. The conventional test used by Waters (1977) was easier in comparison to the stradaptive test than was the conventional test employed by Betz (1976a) for their respective groups of examinees. Thus, the latency results depend on the particular conventional test employed and also on the ability

level of the group to which it is administered. Generally, however, the adaptive test will tend to administer items very near to the ability level of the examinee so that in comparison to the conventional test, this may be a more or less difficult test for a given set of examinees. Although item latency data showed significant differences between stradaptive and conventional tests in the present research, in practical terms the difference in mean total testing time for a 40-item test would be approximately 1 minute.

Although the stradaptive tests took slightly longer to administer, the response pattern information data showed that they provided substantially more information than did the conventional tests. With equal test length to that of the conventional tests, the fixed-length stradaptive test provided measurements that were more than twice as precise as those of the conventional test, on the average. This result can be translated directly into test length savings of more than 50% to attain levels of precision equal to those of the conventional test. Even more precise measurements were obtained by the variable-length stradaptive test, which obtained measurements with mean information more than twice that of the conventional test, while administering almost 50% fewer items. This indicates that a variable-length stradaptive test would require only about 17.5 items to achieve the same average level of measurement precision as the 50-item peaked conventional test. These results are consistent with both earlier live testing and simulation studies demonstrating the measurement superiority of the variable-length stradaptive test (e.g., Thompson, 1980; Vale, 1977; Vale & Weiss, 1975a, 1975b).

In a number of studies (Betz & Weiss, 1976b; Pine, Church, Gialluca & Weiss, 1979; Prestwood & Weiss, 1978), there has been no effect on test anxiety due to feedback. Similarly, in the present study, there was no decrease in mean anxiety as assessed by the Anxiety Scale when students received feedback. It may be that the volunteer experimental subjects did not have enough interest to perform well on the test to become test-anxious. College students taking an experimental test may have a low level of test-taking arousal. The effects of feedback in a motivated context (for example, the classroom) on anxiety may differ from that found in an experimental setting; or it may be that test anxiety is a stable expectation of performance for a person who is fairly resistant to testing conditions, such as type of test or administration of feedback. Performance and anxiety level could possibly be affected by altering the quality of feedback, i.e., by using a relatively easier test adapted to the individual's ability. All students would receive a relatively easier test, and the positive feedback could lead to better test performance and lowered anxiety. It may be that students should be grouped into high- and low-anxiety groups-- if anxiety is a stable person characteristic--as well as high and low ability groups; and the effects of feedback under high and low difficulty testing conditions, feedback conditions, and conventional vs. adaptive conditions should be examined. Adaptive testing may have its own motivating effect, since through subjective KR students may perceive the difficulty of the test differently from conventional tests.

The significant three-way KR · Pacing · Testing Strategy interaction indicates that anxiety interacts in a complex way with testing conditions. It is interesting, however, that when KR was provided, the reported level of anxiety did not vary significantly as a function of testing condition, i.e., students receiving KR reported about the same level of anxiety in each of the testing

conditions. When feedback on performance was withheld, students reported anxiety levels that varied significantly as a function of testing condition, with significant differences occuring only for the conventional test. These data suggest that feedback may standardize testing conditions with respect to test-taking anxiety and that under no feedback conditions, students' motivational reactions to conventional tests are more susceptible to the influence of test administration conditions than are their motivational reactions to adaptive tests.

Research has consistently shown that, unlike anxiety, reported motivation varies with the provision of feedback (Betz & Weiss, 1976b; Pine, Church, Gialluca, & Weiss, 1979; Prestwood & Weiss, 1978). Betz and Weiss (1976b) found a significant Ability Group × KR interaction, which indicated differences in motivation between low-ability and high-ability students attributable to the provision of feedback. In the high-ability group, motivation was higher under KR than under no-KR conditions; whereas in the low-ability group, motivation was lower under the KR than under the no-KR conditions. Postulating that motivation increased when the proportion of positive feedback increased, Prestwood and Weiss (1978) studied the joint effects of provision of KR, test difficulty (proportion of positive feedback), and testing strategy on test performance and on test-taking reactions with students of high ability. The examinees reported higher motivation when KR was provided than when it was not. Thus, higher ability students in both studies reported higher motivation when KR was provided.

In the Prestwood and Weiss (1978) study there was no effect of test difficulty on reported motivation nor was any interaction significant. This would indicate that it is not merely the quality of feedback (positive or negative) that determines motivation to perform well but the examinee's reaction to feedback. Positive or negative reaction to feedback may be determined in part by a history of academic successes or failures. In other words, two examinees may receive the same amount of negative feedback but will react differently because of differences in academic history. This conclusion was partially supported by a finding reported in Pine, Church, Gialluca, and Weiss (1979) that Black examinees were less motivated under KR conditions than White examinees, who were more motivated under KR conditions. Although the study did not control for proportion of positive feedback, it may be that Black students reacted less favorably to feedback than White students due to differences in academic history.

Although the present study did not replicate the positive relationship between KR and reported test-taking motivation, it did indicate that motivation level varied as a result of the interaction of feedback with pacing of item presentation. The interaction indicated that highest motivation was reported under computer-paced KR conditions, whereas lowest motivation was reported under computer-paced no-KR conditions. Such a finding indicates that the effect of KR on motivation is modified by other testing conditions with which feedback is paired. Differences in empirical results dealing with the effect feedback has on test-taking motivation may stem from variation in testing conditions--such as pacing of item administration, test difficulty, and testing strategy--with which feedback has been paired. As with the anxiety interaction, there was evident in the motivation interaction the possible standardizing effect of feedback. Levels of reported motivation were not significantly different between computer- and self-paced conditions when feedback was provided, but statistically significant differences were observed between pacing conditions when feedback was not

provided. Unlike the Motivation Scale scores, however, the significant differences for the Anxiety Scale scores did not occur only on the conventional tests.

Chi-square analysis of post-questionnaire attitude-assessment items indicated that feedback and testing strategies tend to affect student reactions to the testing environment in different ways. As there were no differences among patterns of responding to the psychological reactions items between self- and computer-pacing conditions, this would indicate that pacing, as it was defined in this study, may not be an important testing variable affecting psychological reactions. Feedback, on the other hand, appeared to decrease the reported tension level of the examinee, to foster attempts of the student to try harder on the test, and to respond to each item with the most appropriate answer. Testing strategy had no measurable effect on the motivation or anxiety state of the student; but because the testing strategies differed in difficulty, it did affect students' perceptions of test difficulty. In general, these differences reflect the correct perception that the stradaptive tests were tailored to the ability level of the individual, whereas the conventional test was peaked at the ability level of the student population. Thus, the conventional test was more often perceived to be too difficult or too easy in comparison to the stradaptive test. Chi-square results presented in Table 13 show that patterns of responding differed between students in conventional and stradaptive test strategies on all Difficulty Perception Scale items. In general, students taking the stradaptive test said that they thought the test was a little difficult for someone of their ability and that they were somewhat frustrated by the test difficulty. More students in the conventional test correctly responded in the extreme categories of the difficulty perception items, indicating that the test was perceived as being more often too hard or too easy.

Conclusions

The standardizing effect feedback has both on psychological reactions to testing and on test performance was a finding that occurred repeatedly in this study and one that should be investigated further. This standardizing effect occurred with the Motivation Scale in a KR × Pacing interaction and on the Anxiety Scale in the KR × Pacing × Testing Strategy interaction. In the former interaction, levels of motivation were more similar between computer- and self-paced conditions when feedback was provided than when it was not. Even more striking was the lack of variation in mean Anxiety Scale scores across experimental conditions when feedback was provided. Students in the six experimental conditions, which derived from combinations of the three testing strategies and two pacing conditions, indicated that anxiety level varied widely when feedback was not provided. Research is indicated to detect if feedback has such a standardizing effect when combined with other experimental testing treatments. Furthermore, such research should deal directly with the apparent standardizing effect of KR.

The present study, using high-ability students and the experimental manipulation of testing strategy, KR administration, and pacing of item administration showed no effect of KR on ability estimates or on reported motivation of the students. The expected increase in ability estimates due to the motivating effects of KR was not found. One reason for this might have been due to some interaction between variables that were not experimentally controlled. Important variables that require particular attention in the study of the motivational

effects of KR may be the ability level of the experimental subjects and the difficulty of the test and, thus, the proportion of positive or negative feedback. A study investigating the effects of KR should be implemented under "motivated" conditions so that the experimental test would count toward a grade in a required course. In this way, the true motivating effect might be better assessed on students who are maximally motivated to perform well. Under such conditions KR might increase anxiety to a detrimental degree and might result in poorer test performance.

That KR has been shown to have an effect on performance and reported motivation in several earlier studies (Betz & Weiss, 1976a, 1976b; Pine, Church, Gialluca, & Weiss, 1979; Prestwood & Weiss, 1978) indicates that it is an important testing parameter meriting further investigation. However, particular attention must be paid to the experimental variables with which it is paired and to the ability, test anxiety, and motivation levels of the examinee groups that would be employed in the study.

REFERENCES

Bayroff, A. G. Feasibility of a programmed testing machine (Research Study 64-3). Washington, DC: U.S. Army Personnel Research Office, 1964.

Bejar, I. I., & Weiss, D. J. Computer programs for scoring test data with item characteristic curve models (Research Report 79-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1979.

Bejar, I. I., Weiss, D. J., & Gialluca, K. A. An information comparison of conventional and adaptive tests in the measurement of classroom achievement (Research Report 77-7). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1977.

Betz, N. E., & Weiss, D. J. Simulation studies of two-stage ability testing (Research Report 74-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1974.

Betz, N. E., & Weiss, D. J. Empirical and simulation studies of flexilevel ability testing (Research Report 75-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, July 1975.

Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability test performance (Research Report 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (a)

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1976. (b)

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968, 397-479.

Cronbach, L. J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

Ferguson, R. L., & Hsu, T. The application of item generators for individualizing mathematics testing and instruction (Report 1971/14). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1971.

Liebert, R. M., & Morris, L. W. Cognitive and emotional components of test anxiety: A distinction and some initial data. Psychological Reports, 1967, 20, 975-978.

Locke, E. A., Cartledge, N., & Koeppel, J. Motivational effects of knowledge of results: A goal-setting phenomenon? Psychological Bulletin, 1968, 70, 474-485.

McBride, J. R., & Weiss, D. J. Some properties of a Bayesian adaptive ability testing strategy (Research Report 76-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1976.

Mandler, G., & Sarason, S. B. A study of anxiety and learning. Journal of Abnormal and Social Psychology, 1952, 47,166-173.

Morris, L. W., & Fulmer, R. S. Test anxiety (worry and emotionality) changes during academic testing as a function of feedback and test importance. Journal of Educational Psychology, 1976, 68, 817-824.

Pine, S. M., Church, A. T., Gialluca, K. A., & Weiss, D. J. Effects of computerized adaptive testing on black and white students (Research Report 79-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1979.

Prestwood, J. S., & Weiss, D. J. Accuracy of perceived test-item difficulties (Research Report 77-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, May 1977.

Prestwood, J. S., & Weiss, D. J. The effects of knowledge of results and test difficulty on ability test performance and psychological reactions to testing (Research Report 78-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, 1978.

Strang, H. R., & Rust, J. O. The effects of immediate knowledge of results and task definition on multiple-choice answering. The Journal of Experimental Education, 1973, 42, 77-80.

Thompson, J. G., & Weiss, D. J. Criterion-related validity of adaptive testing studies (Research Report 80-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1980.

Vale, C. D. Strategies of branching through an item pool. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1975.

Vale, C. D., & Weiss, D. J. A study of computer-administered stradaptive ability testing (Research Report 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1975. (a)

Vale, C. D., & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1975. (b)

Vale, C. D., & Weiss, D. J. A comparison of information functions of multiple-choice and free-response vocabulary items (Research Report 77-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, April 1977.

Waters, B. K. An empirical investigation of the stratified adaptive computerized testing model. Applied Psychological Measurement, 1977, 1, 141-152.

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973.

Weiss, D. J. Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1974.

Zontine, P. L., Richards, H. C., & Strang, H. R. Effects of contingent reinforcement on Peabody Picture Vocabulary Test performance. Psychological Reports, 1972, 31, 615-622.

APPENDIX: SUPPLEMENTARY TABLES

Table A
Item Numbers, and Discrimination (a) and Difficulty (b)
Parameter Estimates for the 50-Item Conventional Test

Item Number	a	b	Item Number	a	b
597	.624	-0	329	1.424	.177
382	.856	-.010	208	.743	-.179
292	.610	.012	670	.872	.196
205	.603	-.024	91	1.132	-.197
207	.793	-.035	622	.444	.201
104	.944	.050	52	.844	.205
137	.499	-.056	661	.799	.206
444	.621	.059	667	.719	-.215
209	.870	.067	502	.730	.218
145	.791	.086	272	1.960	.223
503	1.062	-.090	211	.773	-.236
355	.506	.104	37	.860	-.236
365	.877	-.105	645	.674	.242
176	.415	-.106	224	.679	-.257
380	1.822	.115	390	.797	-.257
154	.872	-.124	327	.795	.258
218	.407	-.125	221	.822	-.278
234	.650	-.132	144	.910	.286
161	1.384	.132	568	1.627	.290
56	1.109	.135	369	.788	.295
270	1.223	-.138	318	.526	.310
143	1.036	-.153	50	.694	-.321
599	1.634	.158	307	.699	.325
156	.841	-.166	116	.494	.334
626	.917	.172	128	1.04	-.355

Item Numbers, Discrimination (a) and Difficulty (b) Parameter Estimates for the Variable-Length and Fixed-Length Stradaptive Tests

Item	a	b	Item	a	b	Item	a		Item	a		Item	a		Item	a	
Stratum 1 (26 items)			Stratum 2 (cont.)			Stratum 4 (cont.)			Stratum 5 (cont.)			Stratum 6 (cont.)			Stratum 7 (cont.)		
25	3.00	-2.63	Mean	1.08	-2.05	203	.82	-.48	191	.62	.08	252	.42	.47	259	.37	1.29
28	3.00	-2.63	SD	.65	.21	83	.80	-.49	292	.61	.02	165	.35	.56	333	.35	1.34
42	3.00	-2.63	Stratum 3 (35 items)			535	.77	-.37	205	.60	-.02	372	.35	.56	516	.35	1.12
64	3.00	-2.36	13	1.89	-1.55	183	.73	-.45	318	.53	.31	Mean	1.08	.61	308	.34	1.31
102	3.00	-2.36	190	1.82	-1.44	110	.70	-.54	335	.51	.10	SD	.70	.18	350	.32	1.51
14	2.21	-2.46	191	1.75	-1.26	185	.68	-.68	137	.50	-.06	Stratum 7 (57 items)			Mean	1.54	1.35
11	1.75	-2.58	84	1.70	-1.64	293	.67	-.57	522	.44	.20	120	3.00	1.46	SD	1.05	.17
24	1.75	-2.37	51	1.43	-1.04	235	.66	-.78	176	.42	-.11	140	3.00	1.38	Stratum 8 (29 items)		
99	1.24	-2.67	129	1.27	-1.35	222	.65	-.50	218	.41	-.13	152	3.00	1.61	115	3.00	2.02
124	1.09	-2.64	40	1.24	-1.03	53	.64	-.48	157	.38	-.25	162	3.00	1.25	193	3.00	2.07
65	1.02	-2.71	101	1.17	-1.40	117	.62	-.66	54	.38	.20	174	3.00	1.46	273	3.00	2.14
181	1.02	-2.58	44	1.15	-1.41	112	.61	-.78	354	.33	.15	254	3.00	1.37	319	3.00	2.14
68	1.01	-2.48	199	1.09	-1.09	393	.60	-.39	547	.30	.22	263	3.00	1.47	359	3.00	2.07
105	.98	-2.63	127	1.08	-1.35	58	.59	-.38	Mean	.86	.03	288	3.00	1.26	360	3.00	1.71
198	.80	-2.50	186	1.07	-1.34	546	.56	-.80	SD	.47	.19	294	3.00	1.07	445	3.00	2.07
16	.75	-2.95	90	.94	-1.31	287	.52	-.65	Stratum 6 (45 items)			321	3.00	1.00	564	3.00	2.29
121	.74	-2.82	86	.89	-1.19	588	.47	-.46	114	3.00	.96	337	3.00	1.18	573	3.00	1.86
89	.72	-2.49	173	.88	-1.06	371	.44	-.92	296	3.00	.67	378	3.00	1.48	591	3.00	1.80
17	.72	-2.89	637	.88	-1.02	615	.44	-.86	347	3.00	.49	541	3.00	1.16	609	3.00	2.14
131	.60	-2.58	83	.88	-1.45	155	.40	-.57	264	2.28	.55	562	3.00	1.22	659	3.00	1.95
81	.44	-2.39	285	.84	-1.02	136	.32	-.56	266	2.12	.51	595	3.00	1.58	514	1.16	1.74
151	.44	-2.65	34	.83	-1.58	Mean	.83	-.62	340	1.92	.65	652	3.00	1.66	400	.93	1.68
73	.43	-2.69	227	.81	-1.25	SD	.32	.19	315	1.85	.52	617	2.78	1.17	521	.75	1.70
135	.43	-2.79	640	.78	-1.06	Stratum 5 (48 items)			342	1.59	.54	534	2.52	1.61	260	.71	1.82
507	.39	-2.74	189	.76	-1.19	630	3.00	.28	265	1.57	.55	253	2.32	1.44	504	.64	1.81
201	.31	-2.97	311	.75	-1.43	272	1.96	.22	301	1.38	.47	383	2.11	1.52	533	.63	2.15
Mean	1.30	-2.62	88	.71	-1.33	599	1.63	.16	386	1.25	.54	299	1.77	1.16	577	.61	2.00
SD	.96	.17	232	.67	-1.25	568	1.63	.29	60	1.23	.64	561	1.72	1.42	616	.61	1.76
Stratum 2 (30 items)			322	.67	-1.09	329	1.42	.18	523	1.21	.88	291	1.64	1.35	374	.56	1.99
7	3.00	-2.32	641	.58	-1.40	161	1.38	.13	582	1.20	.35	306	1.32	1.20	403	.54	1.76
71	3.00	-2.32	241	.57	-1.05	270	1.22	-.14	538	1.18	.52	601	1.32	1.10	119	.53	1.73
196	2.13	-1.79	94	.56	-1.02	91	1.13	-.20	526	1.17	.92	572	1.29	1.43	167	.42	2.16
138	1.73	-2.02	108	.54	-1.16	56	1.11	.14	59	1.09	.60	581	1.26	1.21	281	.40	1.79
9	1.45	-2.24	141	.48	-1.21	503	1.06	-.09	651	1.09	.89	217	1.25	1.38	603	.38	1.80
27	1.43	-1.68	214	.48	-1.49	143	1.04	-.15	113	1.06	.68	598	1.08	1.04	531	.35	1.92
70	1.29	-2.24	650	.45	-1.55	104	.94	.05	666	1.00	.85	668	.93	1.49	362	.34	2.06
125	1.24	-1.88	276	.45	-1.53	144	.91	.29	146	.93	.47	168	.91	1.55	511	.32	2.28
22	1.20	-1.97	26	.36	-1.02	365	.88	-.11	551	.90	.34	215	.91	1.07	Mean	1.58	1.94
96	1.13	-1.72	Mean	.92	-1.27	154	.87	-.12	271	.89	.80	304	.89	1.34	SD	1.22	.18
206	1.11	-2.19	SD	.41	.19	670	.87	.20	302	.85	.85	587	.87	1.36	Stratum 9 (16 items)		
158	1.08	-2.00	Stratum 4 (39 items)			209	.87	.07	375	.83	.93	231	.87	1.19	118	3.00	2.41
134*	1.07	-1.94	194	1.79	-.96	37	.86	-.24	111	.82	.94	147	.83	1.47	180	3.00	2.63
126	.96	-2.27	36	1.64	-.79	382	.86	-.01	506	.81	.58	610	.79	1.57	245	3.00	2.63
66	.87	-2.02	87	1.24	-.76	52	.84	.21	633	.71	.47	159	.77	1.56	274	3.00	2.72
80	.86	-2.25	501	1.20	-.55	156	.84	-.17	267	.65	.77	238	.76	1.13	353	3.00	2.51
262	.77	-1.93	109	1.11	-.70	221	.82	-.28	233	.65	.41	367	.72	1.40	381	3.00	2.36
5	.75	-2.16	43	1.11	-.86	390	.80	-.26	139	.61	.79	107	.69	1.59	528	3.00	2.82
184	.73	-2.19	515	1.08	-.71	207	.79	-.04	377	.59	.39	164	.69	1.14	545	3.00	2.82
31	.72	-2.14	128	1.07	-.36	145	.79	.09	258	.57	.64	216	.67	1.40	585	3.00	2.82
63	.69	-2.14	522	1.06	-.39	369	.79	.30	133	.57	.56	397	.65	1.34	604	3.00	2.82
106	.67	-2.01	103	1.06	-1.00	211	.77	-.24	593	.56	.55	342	.63	1.28	228	2.94	2.41
255	.64	-2.18	47	1.04	-.96	208	.74	-.18	590	.54	.62	525	.57	1.51	316	1.16	2.68
202	.62	-2.17	332	.97	-.40	502	.73	.22	629	.53	.42	242	.52	1.57	569	.40	2.40
76	.62	-1.75	130	.95	-.44	50	.69	.32	519	.53	.44	505	.50	1.43	243	.32	2.61
559	.62	-1.68	239	.94	-.71	224	.68	-.26	324	.52	.77	368	.46	1.42	580	.32	2.64
95	.56	-1.71	85	.93	-.67	645	.67	.24	116	.49	.33	213	.43	1.43	357	.31	2.68
82	.54	-2.31	204	.88	-.74	234	.65	-.13	289	.48	.69	576	.43	1.13	Mean	2.22	2.62
93	.52	-2.18	46	.84	-.36	597	.62	-0	549	.43	.35	172	.38	1.36	SD	1.21	.16
643	.49	-2.03	123	.82	-.56	444	.62	.06									

DISTRIBUTION LIST

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-711
NAVTRAEQUIPCEN
Orlando, FL 32813

1 Chief of Naval Education and Training
Liason Office
Air Force Human Resource Laboratory
Flying Training Division
WILLIAMS AFB, AZ 85224

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Henry M. Halff
Department of Psychology,C-009
University of California at San Diego
La Jolla, CA 92093

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 CDR Robert C. Kennedy
Head, Human Performance Sciences
Naval Aerospace Medical Research Lab
Box 29407
New Orleans, LA 70189

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Library
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92138

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Office of the Chief of Naval Operations
Research Development & Studies Branch
(OP-115)
Washington, DC 20350

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland (O3B)
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Arnold Rubenstein
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Dexter Fletcher
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria,VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Attn: PERI-OK
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. Robert Ross
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Vasmer
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commandant
US Army Institute of Administration
Attn: Dr. Sherrill
Ft Benjamin Harrison, IN 46256

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. Ross L. Morgan (AFHRL/LR)
Wright-Patterson AFB
Ohio 45433

1 Research and Measurment Division
Research Branch, AFMPC/MPCYPR
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

1 Dr. Marty Rockway
Technical Director
AFHRL(OT)
Williams AFB, AZ 58224

1 Jack A. Thorp, Maj., USAF
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Dr. William Graham
Testing Directorate
MEPCOM/MEPCT-P
Ft. Sheridan, IL 60037

1 Military Assistant for Training and Personnel Technology
Office of the Under Secretary of Defense for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

1 HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

Civil Govt

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Joseph I. Lipson
SEDR W-638
National Science Foundation
Washington, DC 20550

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev. and Research
National Science Foundation
Washington, DC 20550

1 Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. Werner Birke
DezWPs im Streitkraefteamt
Postfach 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. David G. Bowers
Institute for Social Research
University of Michigan
P. O. Box 1248
Ann Arbor, MI 48106

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

Non Govt

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Charles Myers Library
Livingstone House
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. William E. Coffman
Director, Iowa Testing Programs
334 Lindquist Center
University of Iowa
Iowa City, IA 52242

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Meredith P. Crawford
American Psychological Association
1200 17th Street, N.W.
Washington, DC 20036

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 Dr. L. J. T. Eggenberger
DIRECTORATE OF PERSONNEL APPLIED RESEARCH
NATIONAL DEFENCE HQ
101 COLUMBEL BY DRIVE
OTTAWA, CANADA K1A 0K2

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Chester Harris
School of Education
University of California
Santa Barbara, CA 93106

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

1 Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208

1 Dr. Douglas H. Jones
Rm T-255
Educational Testing Service
Princeton, NJ 08450

1 Professor John A. Keats
University of Newcastle
AUSTRALIA 2308

1 Dr. Mazie Knerr
Litton-Mellonics
Box 1286
Springfield, VA 22151

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 Dr. Michael Levine
Department of Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat
Groningen
NETHERLANDS

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gary Marco
Educational Testing Service
Princeton, NJ 08450

1 Dr. Scott Maxwell
Department of Psychology
University of Houston
Houston, TX 77004

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army Navy Drive
Arlington, VA 22202

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MINRAT M. L. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 1328
D-53 BONN 1, GERMANY

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
4 Hill Hall
Columbia, MO 65211

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20912

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Committee on Cognitive Research
% Dr. Lonnie R. Sherrod
Social Science Research Council
605 Third Avenue
New York, NY 10016

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. Robert Smith
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903

1 Dr. Richard Snow
School of Education
Stanford University
Stanford, CA 94305

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN
THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Psychometric Research Group
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research
Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 DR. THOMAS WALLSTEN
PSYCHOMETRIC LABORATORY
DAVIE HALL 013A
UNIVERSITY OF NORTH CAROL
CHAPEL HILL, NC 27514

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Karl Zinn
Center for research on Learning
and Teaching
University of Michigan
Ann Arbor, MI 48104

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

81-1. Review of Test Theory and Methods. January 1981.
80-4. An Alternate-Forms Reliability and Concurrent Validity Comparison of Bayesian Adaptive and Conventional Ability Tests. December 1980.
80-3. A Comparison of Adaptive, Sequential, and Conventional Testing Strategies for Mastery Decisions. November 1980.
80-2. Criterion-Related Validity of Adaptive Testing Strategies. June 1980.
80-1. Interactive Computer Administration of a Spatial Reasoning Test. April 1980.
Final Report: Computerized Adaptive Performance Evaluation. February 1980.
79-7. Effects of Immediate Knowledge of Results on Achievement Test Performance and Test Dimensionality. January 1980.
79-6. The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models. December 1979.
79-5. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement. November 1979.
79-4. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-3. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.
79-2. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.
79-1. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.
78-5. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.
78-4. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-3. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-2. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.
78-1. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
77-7. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-6. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.
77-5. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-4. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-3. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-2. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-1. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-4. An Empirical Study of Computer-Administered Two-Stage Ability Testing. October 1973.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Computerized Adaptive Testing Laboratory
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, MN 55455

END

DATE
FILMED

5 81

DTIC